



# From Paper to Screen: The Theory and Practice of Digitizing Dictionaries



# Housekeeping

- Attendance
- Computers
- Laptop: if you have it, bring it
- Software to install (I will send you an email)
- Dictionary samples: send to me by email

# Course Topics

- What is a dictionary?
- e-Lexicography and Digital Humanities
- Digitization Workflow
- XML and data modeling
- Text Encoding Initiative
- Working with oXygen XML Editor

# Introductions

- What is your name?
- Where do you come from? What do you study or where do you work?
- How come you're in this Summer School and not on a beach?
- How and where do you use dictionaries? Print? Online?
- Do you have a favorite dictionary? Why?
- What dictionary sample did you bring to work on this week?
- Do you have any specific expectations from this particular course?

# Exercise

- Work in groups of three!
- Define DICTIONARY in your own words. Try not to look it up in a dictionary, but rely on yourself and your partner's knowledge and intuition
- Email definition to [ttasovac@humanistika.org](mailto:ttasovac@humanistika.org) with Subject: Dictionary Definition

# Participants' Definitions

- According to our group, a dictionary is a reference work that shows, in an organised manner, information on the meaning of the general/specialised lexicon of one or more languages. It provides definitions, linguistic and cultural information, as well as the possible uses of that particular word/ expression.

# Participants' Definitions

- Un diccionario es un conjunto de palabras (entradas) definidas segun los conocimientos y los usos.
- Um dicionário é: uma ferramenta, física ou digital, monolingue ou multilingue, geral ou especializada, que descreve e/ou define palavras ou termos.
- A set of words and their meanings, organized according a pre established criteria, and usually associated with a meaning.

# Participants' Definitions

- **Dictionary** is a work that collects, usually in alphabetical order, the vocabulary of a particular language, and which is used to clarify doubts about the meaning of a word, the spelling, and providing other useful information to the users (phonetic, grammatical, pragmatical)

# Participants' Definitions

- Dictionary:
- lexical repository of a selected number of words ordered in a alphabetic order typically with grammatical information; small concise and clear definitions, and examples of usage in discourse to clarify users about the meaning of the words



# What is a dictionary after all?



# **Justices Turning More Frequently to Dictionary, and Not Just for Big Words**

of  
possession & belonging

**Coarctate** (coarcto) to strain, to gather a matter into few words, to shorten.

**Coassation** (coassatio) a planking with boards, a boarding or joyning a floor.

**Coxation** (coaxatio) the croaking of Frogs or Toads. Dr Fearly in his Dipper.

**Coccinean** (coccineus) died into scarlet, or crimson colour.

**Cottible** (collibilis) easie to be sod or baked.

**Cochinchin** (Lat. Coccus or Spa Co.)

**Cuchamel** { a kinde of dust or grain, wherewith to dye the Crimson or Scarlet colour; it is a little worm breeding in a certaine place which they call Holy-Oke, or Dwarf Oke, and is found in Cephalonia and other places, on the leaves whereof there ariseth a tumor, like a blister, which they gather, and rub out of it a certain red dust, that converts (after a while) into worms, which they kill with wine (as is reported) when they begin to quicken. Bac. Nat. Hist.

**Cotter**, is a seal pertaining to the Custom-house. Regist. Orig. fol. 192. a. Also a Scrowle of Parchment, sealed and delivered by the Officers of the Custom-house to Merchants, as a warrant that

rutes of Bread and Ale, that An. 51. H. 3. where you have mention of bread Cocker, Wastel-bread, bread of Trete, and bread of common wheate. Cowel.

**Cockle-stayres** (cockles) a pair of winding stairs. See H. Wor.

**Cocklearp**, pertaining to such fowls. Cocks. Dr. Cocks.

**Cockle** (coccus) a small

in the found of Bow-bell

London, which term came first according to Mins-

Son riding with his Father

try, and being utterly igno-

lant, asked his Father what

the horse doth neigh: ridin-

further, the Son heard a Crow

crow, and said, loth the Crow

neigh too! Hence by way

of jeer he was called Cock-neigh.

A **Cockney**, according to some, is a child that liveth long: But Erasmus takes it for a child wantonly brought up, and calls it in Lat. Mar-

mill-street, was anciently so called.

**Cooton** (coffio) a seething, boiling or digesting.

**Cocive** (coctivus) sodden, easily boiled, soon ripe.

**Code** (codex) a volume containing divers books; more particularly a Volume of the Civil Law so called,

**Codice** (codicil) a small

little book, a Schedule or sup-

plement to a Testament, a

Codicil to a ship, and the

Codicil to a boat tied to the

sid ship. Codicil is now

the name of a Testimony annexed to

the Testament, when any

thing is omitted which the

Testator would add, or any

thing put in, which he would

retract. A Codicil is the same

that a Testament is, but that it

is without an executor. See Swinburn in his Treatise of Testaments, and Wills, p. 1. sect. 5.

**Codinatack** (Fr. Corignac)

conserves of gooseberries

in Dominium virorum

Peremptionem. i. be married,

according to the Ceremony

called Coemption, whereby the

Husband and Wife seemed to

one another. Livie. See

more of this in Godwins An-

thology, chap. de nuptiis, &c.

**Coercibilis** (coercibilis) which may be bridled or restrained.

**Coerterio** (coerterio) restraint, keeping in order and subjec-

tion, and punishment.

**Cocaneous** (cocaneus) which is of the same time and age.

**Coeternal** (coeterinus) of the same eternity, lasting together for eternity, equally eternal.

**Coebala** (from eon and eum) that are of the same age or time.

**Cogitatibe** (cogitativus) musing, pensive, full of thoughts.

**Coffa or Cauphe**, a kind of drink among the Turks and Persians, (and of late introduced among us) which is black, thick and bitter, distilled

@ttasovac

## Coffa or Cauphe, a kind of drink among the Turks and Persians, which is black, thick and bitter, distilled from Berries of that nature, and name, thought good and very wholesome: they say it expels melancholy, purges choler, begets mirth, and an excellent concoction.

— Thomas Blount, *Glossographia* (1661)

# Quirkiness matters

or Down With the Tyranny of Corpus Frequency!

# A dictionary is

- a text
- a tool
- a model of language
- a cultural artifact

# Dictionary as a socio-technical assemblage

- Assemblage (Deleuze and Guatarri, 1987)
  - Ordered system, represents stratification of reason
  - A product of the nomadic spontaneity of language, "bearing a foreign tongue within each language as a whole" (514)

# Dictionary as Text

- A book (OED, Chambers)
- A reference work (Duden, Wiktionary)
- ... containing a “list of words” and their explanations.

# Dictionary as Text

- Specialized genre
- Strict, canonical structure
- Condensed mode of expression

*Le dictionnaire n'est pas seulement un objet, un produit de consommation, défini par des besoins socio-culturels, c'est aussi et surtout un texte, un discours continu et clos.* (Dubois 1970: 35)

# Dictionary as Text

- What does it consist of?
- Basic textual units: entries
- System of pointers showing mutual connections
- Fragmentary, yet it builds a whole ("ein Textganzes")

# Dictionary as Text

*With the dictionary as text, one does not locate the semantic answer; one reads signifiers explicating other signifiers. Reading is a contingent act...Reading involves constructing meaning, not absorbing it. (Frawley 1985: 11)*

# Defining the Dictionary: Book vs. Text?

- What is the difference?
- Specific medium vs. written discourse

# Print Culture

- "typographic fixity" made possible by the printing presses (Eisenstein 1979: I, 116-120)
- Data collection, retrieval and storage, patterns of communication
- Perceived stability and palpability of the book
- Mass reproducibility

# Dictionary as a book (das Wörterbuch, a wordbook)

- Not a terminological accident
- Complex cultural concept which evolved over time

# Dictionary as prototypical (hyper)text

- What is a hypertext?
- a network of nodes and links which, unlike traditional, sequential texts, are meant to be read in a non-linear fashion

# Dictionary as Text

*words are defined in terms of other words, which are then defined in terms of other words, and so on. The circularity of the dictionary is a product of the dictionary as text. (Frawley, 1985, 10)*

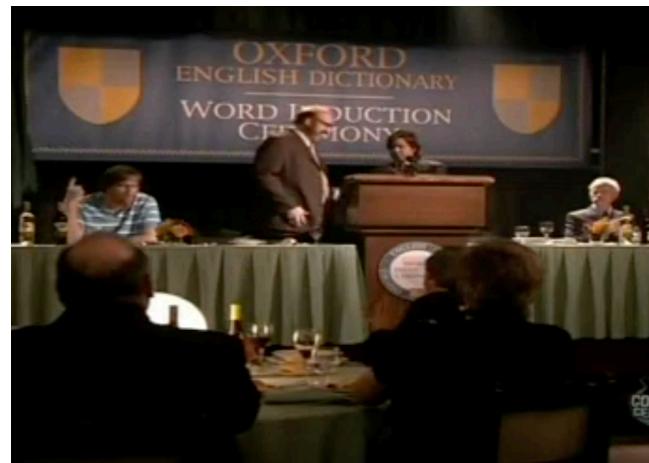
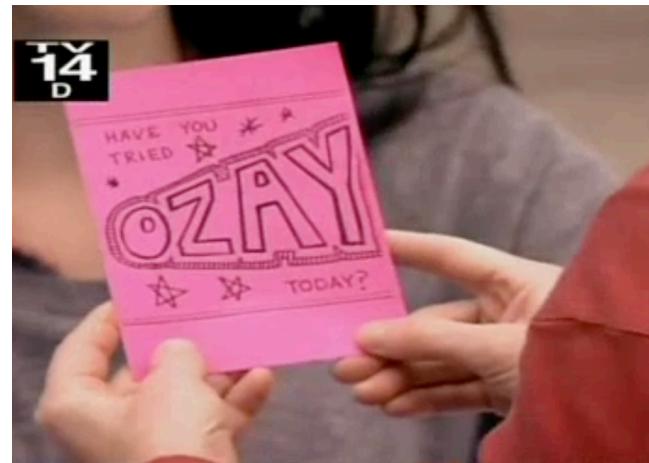
# Lexicographic order as literary inspiration

- Ouvroir de littérature potentielle (Oulipo) : S+7
- Milorad Pavić, *The Dictionary of the Khazars: A Novel Lexicon in 100,000 words*

*The user of this dictionary, however, should not be discouraged by these detailed instructions. He can easily skip all these preliminary notes and read as he eats: reading, he could use his right eye like a fork, and his left eye, like a knife, throwing bones behind his back. And that would be enough. Admittedly, the reader could lose track and get lost among the words of this book, the same way that Massudi, one of the writers of this dictionary, lost his way in other people's dreams and never found his way back.*

(Pavić, translation mine)

# Sarah Silverman Program (2008)



# Dictionary as Model of Language

- Prescriptive vs. descriptive
- language as a unified, delimited system consisting of discrete meaningful units (words) which have basic, fixed meanings (Sargeant 2011)
- Dictionary as an instrument of lexical legitimization

# Corpus vs. dictionary

- Word as a privileged unit of meaning?
- "The corpus contains traces of meaning events; the dictionary contains lists of meaning potentials." (Hanks 2008: 130)

# Corpus vs. dictionary

- "A lexicographical definition, we shall argue, does not in most cases identify a meaning independently existing in actual usage and discovered there by the lexicographer: it is deliberately constructed and allocated by the lexicographer on the basis of materials selected for study, and its allocation will depend on the viewpoint the lexicographer has chosen to adopt." (Harris and Hutton 2007: 78)

# Lexicographic theory

- Scerba (Щерба 1940),
- Zgusta (1972),
- Hausmann (1977)
- Kromann et al. (1984)

# Wiegand: dictionary as a utility product

- Gebrauchsgegenstand
- Satisfies user needs
- Detailed typology
  - Language dictionaries (Sprachwörterbücher)
  - Dictionaries for special purposes (Sachwörterbücher)
  - Encyclopedic dictionaries (Allwörterbücher)

# Function Theory of Lexicography

- Center for Lexicography at the Aarhus School of Business in Denmark
- "A lexicographical function is the satisfaction of the specific types of lexicographically relevant need that may arise in a specific type of potential user in a specific type of extralexicographical situation" (Tarp 2008: 81)."
- Lexicography as a social science which deals with information management, rather than language as such

# Function Theory of Lexicography

- User needs
  - cognitive
  - communicative
  - operative
  - interpretive

# Special case: historical dictionaries

- Function Theory does not deal with historical dictionaries
- Use cases:
  - difficulties in the reception of historical texts
  - difficulties in the production of modern translations;
  - Answer questions about linguistic and cultural traditions (see Reichmann 2012: 54).

# Dictionary as a cultural artifact

- **Shaped by** social values of its time (religion, literature, education, politics, economics and language planning)
- **Shaping** social values of its time and playing an important role in all four stages of language standardization (selection, codification, acceptance and elaboration of a standard)

# The Academy tradition

- Vocabolario degli Accademici della Crusca (1612 ff)
- Dictionnaire de l'Académie française (1687 ff)
- Diccionario de la lengua castellana (1726 ff)

# Dictionary and nation building

- Webster, American Dictionary of the English Language (1828)
- Reforming American spelling in order to “make a difference between the English orthography and the American” — a move which would be of “vast political consequence” because “a national language is a band of national union” (Webster 1789/1991: 87).

# Dictionary logic?

- Order (reason) or madness (dreams?)
- *Dictionnaire critique in Documents*  
(1929-1930), edited by Georges Bataille

# Dictionary logic?

- Order (reason) or madness (dreams?)
- *Dictionnaire critique in Documents*  
(1929-1930), edited by Georges Bataille

# Bataille, *Factory chimney*

- "Giant scarecrows" installing fear and anguish in him as a child "comme la révélation d'un état de hoses violent dans lequel il se trouve pris à partie" [as the revelation of a state of violence for which one bears some responsibility]:
- A cette manière de voir enfantine ou sauvage a été substituée une manière de voir savante qui permet de prendre une cheminée d'usine pour une construction de pierre formant un tuyau destiné à l'évacuation à grande hauteur des fumées, c'est-à-dire pour une abstraction. Or, le seul sens que peut avoir le dictionnaire ici publié est précisément de montrer l'erreur des définitions de ce genre. (Bataille 1929b: 332)
- [These childish or untutored way of seeing is replaced by a knowing vision which allows one to take a factory chimney for a stone construction forming a pipe for the evacuation of smoke high into the air — which is to say, for an abstraction. Now, the only possible reason for the present dictionary is precisely to demonstrate the error of that sort of definition.]

# So, a dictionary, you say?

- Complex socio-cultural and technical assemblage with a long history
- Text, tool, model of language, cultural artifact
- Culturally constructed and culture-constructing
- An anti-dictionary (cf. Bataille) is still a dictionary

# Exercise: prepare short (5-min) presentations

- **Group 1:** Brief overview of monolingual Portuguese dictionaries (print only)
- **Group 2:** Brief overview of monolingual Portuguese dictionaries (electronic)
- **Group 3:** *Saudade*: the limits of dictionary definitions and cultural understanding
- **Group 4:** Pick a word and show how it's treated differently in a Portuguese and a Brazilian dictionary. What can we learn about the dictionaries' editorial policies from this example.
- **Group 5:** Pick a word or two from Dicionário inFormal (<https://www.dicionarioinformal.com.br>), analyze them and compare them to more traditional dictionaries.
- **Group 6:** Pick one word from Urban Dictionary and one from Dicionário inFormal. What can the differences between these two words and how they are treated tell us about these two dictionaries?